DOCUMENT RESUME

ED 163 058 TM 008 110

AUTHOR	Stufflebeam, Daniel I.
TITLE	A Response to the Michigan Education Department's Defense of Their Accountability System. Paper #1 in Occasional Paper Series.
INSTITUTION	Western Michigan Univ., Kalamazoo. School of Education.
SPONS AGENCY	Carnegie Corp. of New York, N.Y.
PUB DATE	Aug 74
NOTE	36p.
AVAILABLE FROM	Mary Anne Bunda, The Evaluation Center, College of Education, Western Michigan University, Kalamazoo, Michigan 49008 ($2.00)
EDRS PRICE	MF-$0.83 Plus Postage. HC Not Available from EDRS.
DESCRIPTORS	Academic Achievement; Criterion Referenced Tests; *Educational Accountability; *Educational Assessment; Educational Objectives; Elementary Education; *Evaluation Methods; Evaluators; *Program Evaluation; *State Programs; Testing Programs
IDENTIFIERS	Meta Evaluation; Michigan; *Michigan Accountability System; *Michigan Educational Assessment Program

ABSTRACT

The Michigan accountability system, based on statewide objectives-referenced testing at two grade levels generated considerable controversy, including this response to criticisms of an evaluation of the program by a panel of three persons. The contract under which the evaluation was performed is quoted to show that the authors of the report were independent and not biased against the practice of accountability systems. The findings of the report are summarized, including the judgments that Michigan's six-step model was endorsed, but that the promise of the model had not been fulfilled. It judged that goals were unclear and somewhat redundant; that untenable claims had been made concerning the development of objectives; that the objectives should not be published in book form; that the scheme for tying state funding to test score results was potentially harmful; and that little had been done regarding two steps of the program--development of local evaluation systems and influence on state and local decision making. Criticisms are made of the state's response to the evaluation reports and to other criticisms. It is concluded that professional educators should participate in the accountability movement. (CTM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED163058

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

A Response to the Michigan Education Department's Defense of Their Accountability System

by
Daniel L. Stufflebeam

August 1974

EVALUATION CENTER

COLLEGE OF EDUCATION
WESTERN MICHIGAN UNIVERSITY
KALAMAZOO, MICHIGAN 49008

"PERMISSION TO REPRODUCE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY
M. A. Buzda
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

Paper #1
in
Occasional Paper Series

IM008 110

WESTERN MICHIGAN UNIVERSITY
KALAMAZOO, MICHIGAN 49008

OCCASIONAL PAPER SERIES

The Occasional Paper Series is published by the Evaluation Center, College of Education, Western Michigan University. It provides information about new developments in evaluation to various educational audiences.

For information about previous publications in the series, see the inside of the back cover of this document.

All correspondence in reference to the Occasional Paper Series should be addressed to:

Information Officer
Evaluation Center
College of Education
Western Michigan University
Kalamazoo, Michigan 49008

EVALUATION CENTER

Daniel L. Stufflebeam
Director

Mary Anne Bunda

James R. Sanders

ABOUT THE CENTER

The Evaluation Center is committed to advancing the science and practice of evaluation. To serve its purpose the Center has developed components in research, development, instruction, dissemination, and service. Over the past ten years the staff of the Center has had significant experience in these areas working with school districts, state departments of education, foundations, and federal agencies.

The *Research Component* is designed to develop and test evaluation theory. Particular thrusts are testing the effects of school-based evaluation offices on school practice, and determining the special evaluation skill and knowledge profiles required by various groups in education. The *Development Component* is designed to work on ways and means of implementing evaluation theory. Included are techniques for meeting special evaluation objectives, new data collection instruments, and special procedures for communicating evaluative findings and relating them to practical problems. The *Instructional Component* is designed to provide in-service training in evaluation to various local, state, and national groups. While the Center is not a credit-granting agency, it collaborates with instructional units throughout Western Michigan University for the development and implementation of pre-service evaluation curricula. Emphasis in this work is at the doctors and masters levels. The Center serves as an instructional laboratory wherein students majoring and minoring in evaluation can gain practical evaluation experience. The Center is also committed to the development of novel and needed evaluation training materials and strategies. The *Dissemination Component* is designed to communicate research findings, new ideas, and materials to the education profession. *The Service Component* is designed to assist agencies in meeting their evaluation needs through two primary services. These are the design and implementation of evaluation studies and the development and installation of agency evaluation systems.

The staff of the Center has been involved in conceptualizing the CIPP Evaluation Model, researching the Item Sampling technique, developing the Advocate Team Technique, developing over 100 standardized tests, designing and conducting many in-service evaluation programs in evaluation. Other accomplishments include authoring the PDK book, EDUCATIONAL EVALUATION AND DECISION MAKING, and numerous journal articles and papers, designing and conducting many evaluation projects, and installing evaluation systems in schools, government agencies, and research agencies. Overall, the Center staff has directed externally funded projects in evaluation totaling more than four million dollars.

ONCE UPON A TIME the principals in a large school system proposed to the superintendent that all the teachers take an annual examination. The information thus gained would supplement the principals' classroom visits (which they were usually too busy to make), and would provide objective data about the qualifications of everyone in the system. It might even show the aptitude of teachers for higher responsibilities.

The superintendent found the suggestion excellent, and promised to present it to the board of education. Indeed, he liked it so well that he decided to expand the plan to include the principals, too, along with other administrative persons. The principals had little objection to this last feature, but did oppose an examination for themselves, asserting that the creative and flexible nature of their work was such that an examination would not give a full and fair picture. And besides, they were willing to rely upon the judgment of the superintendent and the central office staff who had selected them in the first place.

But the superintendent was still enthusiastic about his idea, and proposed it one night to the board of education. They were even more enthusiastic than he, and endorsed it wholeheartedly. They felt scores could be posted and persons getting the highest scores could be formally honored with a dinner and a plaque. They had only one change to suggest. They would like to request that the superintendent be examined also, to give an inspiring example to all, and a challenge to all in the system to equal his score.

At this point the superintendent praised the wisdom of the board, but warned that refining the proposal, preparing the examinations, and implementing the plan would take some time. Indeed it did, more time than anyone was able to give it.[1]

This story was probably written by a frustrated evaluator who had become cynical about educators' sincerity in supporting evaluation work. As the story suggests, educators are often gratuitous in their support of evaluation--favoring it when applied to somebody else's efforts, but resisting any and all attempts to evaluate their own work.

Fortunately, this situation is not as true today as it was when the above story was written. In 1965 the Congress began requiring that school districts evaluate their use of federal funds. Since then, school

1

districts, universities, federal education agencies, and state education departments have greatly increased their evaluation activities; also there has emerged a nationwide educational accountability movement. This increase in evaluation and accountability work has been nowhere in greater evidence than in Michigan.

Particularly Michigan has emerged as a leader in its development of a statewide educational accountability system. Started in the late 60's, this system is based on a six-step accountability model and has received hundreds of thousands of dollars in support from the Michigan legislature. Annual reports from the system have rank ordered Michigan school districts based on the reading and mathematics test performance of their fourth and seventh grade students. Through its Chapter 3 compensatory education program the system has involved many urban school districts in performance contracting. Recently the system introduced objectives-referenced testing on a statewide basis. Also there are plans for greatly expanding the state accountability system. Clearly, Michigan educators are heavily engaged in evaluation and accountability work.

However, this innovative effort has met with mixed reactions. The legislature has obviously been satisfied as is evidenced by their continuing financial support of the system. Also, the system has been favorably described in the public press and the professional literature. Conversely, the system has also received a great amount of criticism from Michigan professional educators. They have charged that the six-step model is not sound and that the state has done a poor job of implementing statewide accountability activities. Such mixed reports about massive, innovative efforts are not unusual; and they are not

undesirable, for they motivate efforts to determine whether an innovation is worthy of continuation, and, if so, how it might be improved.

It was for just such reasons that Ernest House, Wendell Rivers, and I recently evaluated the Michigan accountability program.[2] Specifically we were engaged during the first three months of 1974 by the National Education Association (NEA) and the Michigan Education Association (MEA) to investigate the educational soundness and utility for Michigan of the state's six-step accountability program. We gathered data about the system through reviewing pertinent documents, conducting interviews, and holding public hearings. We reviewed these data and developed joint conclusions about the system. We then wrote and disseminated a panel report which was both laudatory and critical of the Michigan effort. We have documented our assertions more thoroughly by assembling a technical report of supporting data and procedures of investigation.[3] In essence we said the six-step model was reasonable but that the state had done a poor job of implementing it.

Our report evoked a number of reactions on both national and state levels. The Phi Delta Kappan[4] characterized the study as exhaustive and indicative that the Michigan accountability model won't work. Dr. Helen Wise,[5] President of NEA said, "The findings . . . confirm some of our worst suspicions that the implementation of accountability systems is counterproductive." In a similar vein, but not in direct response to our study, State Superintendent John Porter was quoted in the June 5, 1974 issue of Report on Education Research[6] as expressing pessimism about the future of educational accountability in Michigan, saying that the

Michigan model "can't succeed in the long run, and that teacher opposition will eventually doom the plan." In contrast with these negative and pessimistic statements, the Michigan Department of Education[7] issued a document that criticized our panel report and responded in depth to its recommendations.

When Dr. Taylor invited me to speak to you he asked that I discuss our panel's recommendations and the Michigan Education Department's response to those recommendations. I accepted this charge, because--as a Michigan educator and specialist in evaluation--I believe I should do what I can to insure that our state accountability system will be a positive force in serving Michigan's educational needs.

Particularly I believe the report that House, Rivers and I developed contains some important issues that merit serious discussion by Michigan's educators. Also I feel I can contribute further to this discussion by providing my views on the Michigan Education Department's published reactions to the panel report. Thus I am pleased to participate with you in a constructive discussion of how Michigan's educational accountability system might be improved.

It should be noted that I will present my own thoughts and not necessarily those of any other party. Whereas the Michigan Department of Education supplied me with much information about their system, in no way do I wish to imply that they would endorse or reject the contents of this paper. Likewise, I am not a spokesman for the MEA or the NEA. While they financed the study that House, Rivers, and I conducted, we retained complete independence in writing, editing, and releasing our report. Finally, I am obviously indebted to House and Rivers for their

contributions to our joint study, but I have not consulted them in the preparation of this manuscript, nor have I restricted myself to the contents of our joint report. Overall, I wish to acknowledge that many persons and groups have contributed to my thinking, but I alone accept responsibility for the contents of this paper.

I

The House-Rivers-Stufflebeam Assessment of the Michigan Accountability System

The study that House, Rivers, and I conducted was motivated by both national and state interests in the Michigan accountability system. Since Michigan has assumed leadership among the states in exploring and applying accountability concepts, it is not surprising that other states, that are considering or already operating statewide accountability programs, are interested in learning as much as they can about the successes and failures of the Michigan program. I believe it was for this reason that the NEA commissioned House, Rivers and me to evaluate the Michigan system. The state-level interest in the Michigan accountability program needs no explanation to educators in Michigan; for the six-step accountability program has been the center of controversy in the state ever since the Department of Education began publishing lists of school districts, rank ordered on the basis of their students' test scores.[8] Because of this controversy, it is easy to understand why the MEA joined with the NEA to sponsor the study that House, Rivers, and I conducted.

Why NEA and MEA chose me to participate with House and Rivers in the study is still a mystery to me. House, Rivers, and I had not worked as a team in the past. I had not worked for MEA or NEA in the past. Whereas, it is alledged that NEA and MEA oppose accountability programs, I--at least--have favored and been involved in their development. For example, I directed a project at Ohio State University which prepared and assessed three alternative designs for a statewide accountability system for Ohio.[9]

If NEA and MEA wanted someone with neutral predilections about Michigan's accountability I would not have been a good choice, since I have often remarked publicly about the strong talent pool in Michigan's accountability program. I can only conclude that I was chosen along with Rivers and House because the sponsors expected that we would do a competent job and because they judged that our reputations in work alligned with accountability are respected by members of the intended audience for the study.

However, I realized early that our study would be controversial, that it would be difficult to conduct in an independent fashion, and that in any case our objectivity likely would be questioned by whomever would oppose our findings. Before agreeing to participate in the study, I met with members of MEA and NEA, with Rivers and House, and with a representative of the Michigan Department of Education to determine for myself whether our team could conduct an independent study and whether we would have access to all the requisite data. After much discussion I became convinced that House, Rivers, and I could conduct an independent assessment and that we would have access to all pertinent available information.

An advance written set of working agreements to govern our study was adopted by Rivers, House, MEA, NEA, and myself. This was the basis for my agreeing to participate in the study. The advance working agreements were included as an appendix in our panel report. Unfortunately, those who have questioned our independence in conducting the study have systematically omitted any mention of the written contract that governed our work and guaranteed its independence. In case our critics failed to read the appendix I have decided to list the working agreements in the

body of this paper. Hopefully, these will allay concerns about the independence of our study; also they may serve as a useful exemplar for those who may in the future evaluate state accountability systems.

The ten advance working agreements were as follows:

1. Charge

The external evaluation panel consisting of Ernest House, Wendell Rivers, and Daniel Stufflebeam have been engaged by the Michigan Education Association and the National Education Association to evaluate the educational soundness and utility for Michigan of the Michigan Accountability Model with a particular focus on the assessment component.

2. Audiences (in priority order)

a. NEA/MEA
b. Decision makers in Michigan's educational system (State Board of Education and State Department of Education)
c. The media (the public)
d. Consumers (parents, PTA, the public, etc.)
e. Technical persons (especially in the area of educational measurement).

3. Report/editing

The panel will be solely in charge of developing and editing its final report. NEA/MEA may write and disseminate any separate statement (such as an endorsement, a rebuttal, a commentary, or a descriptive piece). It is understood that the panel's report is to be as short and direct as possible and to be designed to communicate with the audiences designated for the report.

4. Dissemination

The external panel has the right to release its report to any members of the target audiences or other persons following the completion of the report. The panel's release of the report will imply no MEA/NEA endorsement. Further MEA/NEA may choose to endorse or not endorse the report depending on their judgment of the quality and appropriateness of the report. Should MEA/NEA decide to disseminate their own document describing the report their document will be identified as their own and not that of the committee. Only the committee's final report as edited by the committee will be distributed with the names of the committee on it.

5. Format of the Report

The following items were identified as desirable ingredients for the panel's final report:

a. citation of the agreements between the review panel and MEA/NEA.
b. presentation of the major findings.
c. presentation of minority opinions, if any.

6. Questions to be Addressed in the Report

Specific questions to be addressed will include:

a. validity and reliability of criterion-referenced tests.
b. use of tests to evaluate staff.
c. merit of the objectives on which Michigan assessment is based.
d. involvement of teachers in developing both objectives and tests.
e. the panel's recommendations for change and further study.
f. comments about the balance of the state effort and appropriateness of expanding the scope of assessment especially given cost factors associated with the projections for improving or expanding Michigan assessment.
g. quality of planning in the Michigan Accountability Program.
h. cost benefit projections for the program.
i. value of Michigan assessment outcomes and reports for different levels of audiences in Michigan.
j. problems of bias in the Michigan Accountability Program.

7. Resources (budget) to Support the Evaluation

Sufficient resources will be made available by MEA/NEA to the external review panel to support eight days of work per panelist to work on evaluation, whatever secretarial support is needed in conducting the evaluation and whatever materials and equipment are needed in the Lansing hearings. It is understood that if any of the panelists need to make long distance telephone calls in collecting opinions about the program from people in Michigan that the panelists will be reimbursed for such expenses provided that an accurate and complete report is made of the purpose of the phone call and who was contacted.

8. Delivery Schedule

The panel is to deliver its final report on March 1 or as soon thereafter as is practicable.

9. Access to Data

It is understood that the Michigan Department of Education will make available to the panel any and all data and reports required by the panel to do the job. This, of course, is restricted to those data and reports that are now available to the Michigan Department of Education regarding Michigan Accountability.

10. Procedures

Pursuant to the above conditions the external three man panel will have control over the evaluation process that it must implement to responsibly respond to the charge to which it has agreed. In accordance with this position the panel has agreed to implement the following general process. Private interviews and hearings will be conducted solely by the panel with representatives of the Michigan Department of Education, representatives of MEA/NEA, representatives of selected groups (teachers, administrators, board members, and educational action groups). The panel will also review documents made available to it by MEA/NEA and the Michigan Department of Education. Finally the panel will conduct a hearing to obtain additional information concerning issues identified by the panel in the course of interviewing various client groups and studying various documents.

In addition to these ten working agreements, House, Rivers, and I found that our efforts were unified through a common view of the importance of accountability. We agreed that accountability should be practiced at all levels of education. We agreed it should serve both to prove and to improve the quality of education. We agreed that different conceptions of educational accountability need to be tested under field conditions and that experimental efforts in accountability should be critically examined prior to widespread implementation. Hence we welcomed the charge to participate, through the role of critic, in an effort to advance the practice of educational accountability in Michigan.

In completing our study we were almost completely successful in implementing our ten point contract. The exception was that we were not able to get cost/benefit data about the Michigan system. In large

measure our success in carrying out our study design was due to the good cooperation we received from all who were asked to provide information for the study. MEA and the Michigan Department of Education supplied us with a large number of documents. We heard more than thirty hours of direct testimony presented by persons representing various levels of Michigan's educational system. And we received and studied a considerable amount of specially prepared written testimony. After amassing such a large amount of information our task became to study it, evaluate it, organize it, and synthesize it into a report for public consumption.

I believe that we were at least partially successful in developing a credible and readable document. Evidence for this is that our report has so far been published in Teacher's Voice[10] and was the cover article in the last issue of the Phi Delta Kappan[11]. Apparently our report has stimulated considerable public interest both on state and national levels. Hopefully it will prove useful to educators who are charged with developing and implementing accountability systems. In addition to the general panel report, a detailed technical report of the findings was developed and is available from the MEA[12]. Given this background, I next turn to the substance of our report.

The Findings of the House/Rivers/Stufflebeam Study

Overall our study findings were mixed. We acknowledged that the Michigan accountability staff is as competent as any similar group in the nation and that the six-step accountability model is logical and useful as a means of communicating about accountability work. However, we reported serious reservations about the state staff's implementation of the model and about various claims they have made for their work.

Concerning the six-step model we reported the following positive points:

"1. Involving persons from throughout the state in defining common goals is a useful way of focusing communication about educational accountability.

2. Translating common goals into objectives potentially provides a broad base of important variables for assessing needs in Michigan's schools.

3. Assessing needs in relation to objectives derived from the common goals should provide information to state and local-level decision makers to help them determine priorities for a variety of needed change efforts.

4. Testing alternative delivery systems should assist the state to develop a research base for assisting schools to adopt innovative strategies that will serve high priority needs.

5. Fostering the development of local evaluation capability should assist the schools to assess local needs; to design, implement, and assess their innovative efforts; and to evaluate their personnel on fair bases.

6. Using feedback from the accountability system to guide state and local educational policy should assist school districts and the state department to fulfill their leadership roles in education.

Thus, the judgment of our panel about the six-step model was positive. In particular I see in the model a reasonable set of guidelines for the Department of Education and schools to follow in practicing accountability within a utilitarian framework. Personally, I would prefer a model that places greater emphasis on deriving objectives that reflect empirically-determined student needs. But, on balance,

I endorse Michigan's six-step accountability model as one alternative approach to accountability that educators should consider.

Implementation of the Model

If our judgment of the model was positive then why has our report been labeled by some as a highly critical indictment of the Michigan accountability system? Simply because we charged that the promise of the model has not been fulfilled. As we said in our report "Our reservations about the model are not with its rhetoric but its implementation. While the state has made some desirable progress in implementing the model, our panel found that a number of activities have not been consistent with the intent of the model and have in fact been counterproductive." In assessing the model's implementation we reported reservations about each of the model's six steps.

The common goals, we said, are unclear and contain redundancies. Also we noted there is no ongoing review of the common goals and provision for updating them. We acknowledged that these criticisms are not crucial, since they pertain to the technical as against the philosophical qualities of the goals. However, because of their functional importance in determining objectives, we recommended that the common goals be made as clear as possible and that procedures be instituted for periodically reviewing the updating the goals.

Regarding the model's second step we acknowledged the state staff has secured wide involvement of educators in translating the common goals into objectives. However, we called the state staff to task for making exaggerated and untenable claims for the results of their objectives-development work.

The testimony presented to our panel by the state staff noted:

> The performance objectives in reading and mathematics and in the other areas should be viewed as a consensus among educators at all levels of the educational system and in all regions of the state as to the minimum behaviors that students should be able to demonstrate at selected levels of the educational continuum.

However, the evidence we obtained does not support claims that the objectives represent a consensus of even a representative sample of educators, or that the objectives are in any practical sense minimal. On the contrary there is substantial evidence that the objectives were chosen by a non-representative sample and are too stringent to be considered minimal for the designated grade levels.

Personally, I do not think that the objectives need to represent a consensus, nor that they have to be truly minimal. In fact I think that satisfying these conditions will be about as easy as finding the Holy Grail.

What the panel objected to is what we considered to be false and misleading claims about the objectives. We recommended that the Michigan Department of Education drop these claims and instead urge caution in the use of and interpretation of test results related to the objectives.

Further, we urged that the state staff abandon its recently announced plans to publish a book of objectives for parents. The reported hopes of state department officials that the book will provide a handy reference for parents to check up on their children's progress in school seems unfounded. Based on the state's performance so far in trying to choose minimal performance standards, it is more likely that such a book will lead parents to develop faulty assumptions concerning what their children are being taught and unrealistic expectations concerning what their children

should be achieving at given grade levels. Worse, the panel noted that the book could lead to a state-controlled, monolithic curriculum.

Quite obviously, the prospects for misuse and misinterpretation of the state objectives are not unlikely, and the possible consequences are not trivial. For these reasons our report urged Michigan educators to require that state leadership personnel act thoughtfully and responsibly in describing and using the state objectives.

It was in the model's assessment component that the panel found the most serious implementation breakdown. The model's promise of providing ongoing needs assessment in relation to the full scope of the common goals has not been pursued. Instead, attention has been limited mainly to reading and mathematics test performance at two grade levels. To no constructive purpose, school districts have been ranked on norm-referenced tests. Objectives-referenced tests have recently been put into full-scale use before being validated. All pupils are being tested when there are no compelling technical or utility reasons for this.

Overall, there is a present danger that the weakness of the assessment component may undermine the total structure of the accountability model. Among the recommendations we made were that the activities be expanded to more fully treat the full scope of the common goals, that the objectives-referenced testing be placed on a voluntary basis and that every-pupil testing be abandoned in favor of a more efficient matrix sampling plan.

Our panel had mixed reactions about the implementation of step 4 of the model. On the positive side, we supported the state-sponsored research and development work being conducted to identify and analyze alternative educational practices. We also commended the state for its concentration of funds on the basic skills problems of disadvantaged children.

However, the panel expressed serious reservations about the implementation of the Chapter 3 Program. Our judgment was that this program is potentially harmful to education in Michigan, in tying money to test scores.

The Chapter 3 strategy is well known to Michigan educators. School districts are told that they will be rewarded if their poor achievers attain minimal standards, determined by the districts and the state, in reading and mathematics. Since, according to the strategy, the districts are not given the funds until and unless the students meet the standards, the implication is that educators are not doing their work well, but that they can and will improve their performance under the promise of financial rewards. The panel's judgment was that this is a gross misinterpretation of the problems in educating disadvantaged children; and that the inherent implications about the professionalism of Michigan educators are wrong and demeaning. In actuality, in the first two years of the program districts were given their Chapter 3 allocations irrespective of their students test performance[13]. Also recently announced plans[14] indicate that, in the future, the penalties to districts that do not meet the student performance standards will be reduced. These items seem to support our contention that Chapter 3 is desirable in concentrating funds on the basic skills needs of disadvantaged children, but that the original funding strategy is faulty.

We also found measurement and statistical problems in the Chapter 3 strategy. It seems certain that financial rewards to school districts are often given or withheld on the basis of measurement errors. This is especially likely because of the use of gain scores and because the students being tested are at the bottom of the state distribution. In regard to the outcomes of Chapter 3 activities, we found dubious the claims -- such

as appeared in the December 11, 1973 issue of the New York Times[15] -- that Michigan's Chapter 3 Program has produced real gains in achievement for disadvantaged children. While we hope this claim is true, it could be established only through the conduct of a rigorous field experiment.

Overall, the panel supported the emphasis being given to improving education for disadvantaged children. But we recommended that the state abandon its practice of rewarding school districts for good test performance of their disadvantaged students, and we urged that the state properly qualify its claims of success for the program until there is defensible evidence to support such claims.

Our panel had little to say regarding Step 5 of the model, since there has been comparatively little activity in assisting school districts to develop their own evaluation and accountability systems. I happen to believe that the major payoff of the six-step accountability model depends most basically on the successful implementation of Step 5.

Evaluation is an essential ingredient in identifying and solving problems. The majority of educational problems occur at the local school district level. Therefore, I wish to underscore our panel's recommendation that the state department greatly expand their activities in implementing Step 5 of the accountability model. And I would add that the Saginaw evaluation system is one that other districts could profitably review as they plan for the installation of their own evaluation programs.

Finally, I consider the implementation of Step 6 of the accountability model. From my perception our panel tried hard -- through our hearings, our interviews and our study of documents -- to find evidence that the Michigan Accountability System has influenced state and local decisions about education. We found little such evidence.

Instead, we found that decisions to be served have not been clarified

and there is no ongoing procedure for determining state and local information requirements that should be served by the accountability system. Neither did we find evidence that the governor, legislators, or state board members have used Michigan assessment information to shape educational policy for the state. Particularly, we found no defense for the position that testing all pupils at specified grade levels on all test items in reading and mathematics provided vital information to any group in the state. Considering the great cost of testing all pupils -- especially to the pupils and their teachers -- we urged that this practice be abandoned until there is clear cause for it.

On this topic of utility our panel reviewed several draft versions of longrange plans for the Michigan program that reportedly had been written by Frank Womer. We found in his analysis of possible goals for state assessment a valuable focus for clarifying what services should be provided to what groups. We also endorsed his suggestions that sampling as opposed to every-pupil testing may be sufficient to meet the purposes of state accountability.

This completes my review of the House/Rivers/Stufflebeam evaluation of the Michigan Accoun[illegible]ram. I have summarized some of the main points that [illegible]eport and hopefully have captured our overall judgm[illegible]. Basically, I've noted that the Michigan accoun[illegible]ealing conceptual scheme, but has been poorly impl[illegible], wi[illegible]ost serious weaknesses appearing in the assessment and Chapter 3 activities. For those who are interested in greater detail than I have provided I suggest that you obtain and study a copy of the complete report that House, Rivers and I submitted. It is available from the Michigan Education Association.

II

State Responses to the House/Rivers/Stufflebeam Recommendations

Since this report was released in March, the staff of the State Department of Education has made a number of responses. These especially include written reactions to our report. Also, the state staff has implemented certain steps that if not motivated by our report are at least criticisms. In this second main part of my paper I will review and discuss both the written reactions and the pertinent actions of the state staff.

The Immediate Response

The almost immediate response of the state staff was unofficial, defensive and emotional. In about fourteen Michigan newspapers[16] Phillip Kearney, the state's Associate Superintendent for Research and Administration, was quoted as saying the House/Rivers/Stufflebeam report was prepared by poor researchers and contained bad, rotten, lousy research. This tirade hardly created a basis for constructive communication between our panel and the state staff.

The Substantive Response

However, John Porter, the Superintendent of Public Instruction, released a substantive response on behalf of the state staff[17]. The staff response does offer a basis for constructive dialogue to advance the practice of educational accountability in Michigan. Point by point it addresses our judgments and recommendations And it describes further actions that the state staff intends to pursue in an effort to improve the Michigan Accountability Program.

In introducing the staff response Superintendent John Porter said ". . . of the nine recommendations made by the Panel, six of them have the full support of the staff, and appropriate modifications will be made with those six recommendations."

The nine recommendations referenced by Porter were among those contained in our report. In studying the state's response it was difficult for me to tell which recommendations the state staff actually endorsed, since the acceptance or rejection of a recommendation in the text of the report was sometimes contradicted by the detailed responses to our recommendations that appeared in an appendix to the state report. Perhaps the main text and appendix of the state report were written by different authors. In any case I shall consider both parts of the report as I discuss the staff response to our recommendations.

The contradictory nature of the staff response is apparent in their reaction to the panel's first recommendation. That recommendation is that the Department modify the claim that the selected objectives are minimal and represent a statewide consensus. In the text, the Department disagreed and suggested that ". . . the panel members were confused as to the nature of minimal objectives and their derivation. . .". But in the appendix the Department admitted that ". . . no one at this time can be certain of the minimal nature of these objectives. . ."; the appendix version of the response went on to announce that ". . .the staff will contract for a survey to be undertaken to verify if, indeed, the various published objectives represent a statewide consensus among the education profession." Also, Robert Huyser[18] in a March 11 memorandum to school district personnel urged caution in interpreting the state assessment results and the minimal standards associated with these. Huyser's cautions were similar to those that

Frank Womer had issued in his January 1974 memo to Michigan educators[19]. [21] Thus, in response to the state staff's charge that our panel was confused about the state's intended meaning of minimal objectives, I acknowledge this and seem to be in good company. If I read the staff response correctly its authors also are confused about what they have under the label of "minimal and consensual objectives" and in effect have modified their claims about the objectives. Overall, I believe that a common core of objectives that represent a consensus is a holy grail, that the department hasn't found it, and that until they do they shouldn't say they have.

Next I consider the second objective: that the Department abandon its plans to publish a book of objectives for parents. The staff response clearly rejected this one -- both in the text and in the appendix. But there was acknowledgment that this area needs to be studied.

I found the argument that the "book is a good way to involve citizens in the educational process" as not convincing. For the basic contents of the book would be the so-called consensual and minimal objectives, and their validity is obviously open to question. Basically I will believe, until shown otherwise, that educational needs among students, schools, and districts are so highly variable that no one set of objectives can adequately respond to them. Hence, in my opinion the effort to present a book of objectives that applies equally to all students, schools and districts is doomed to failure from the beginning. I repeat: the state ought to abandon or reconceptualize their attempt to provide a book of objectives to parents.

Regarding the third recommendation -- that the Department abandon its practice of rewarding school districts for good test performance of their disadvantaged students -- I am again confused about the state's response. In the text there is agreement that changes are needed in the

Chapter 3 program, but the needed changes are not specified. The statement in the appendix mainly argues, with summary statistics, that the Chapter 3 program has been singularly successful in raising the achievement levels of disadvantaged children in Michigan. While I hope this claim is true the supporting data are hardly sufficient to establish its validity.

The appendix response also charges that the third " . . . recommendation represents a blatant disregard for the well-being of the many poor White, Black, Indian and Chicano pupils that are underachieving in public schools". This is an irresponsible statement, and I strongly resent it. Anyone who has read our report knows that House, Rivers and I commended the state for its concentration of funds on the problems of disadvantaged children and called for a continuation of this support. As the state staff knows, this is an entirely separate issue from the one supporting the education of poor children only after they and their teachers through some sort of bootstrap operation have raised test score performance. What I recommend is that schools with high concentrations of disadvantaged children be provided money before the fact so that this money can be used constructively in the process of improving instruction. I submit that this recommendation represents a concern and not a disregard for the needs of disadvantaged children, and I would add that it offers realistic support for those teachers and administrators who have the difficult task of educating disadvantaged children.

In regard to recommendation 4 -- that the Department expand its activities in implementing Step 5 of the Accountability Model, the staff response and I are in essential agreement. The state staff reportedly is conducting a survey of local and intermediate school district resources and capabilities in evaluation. Incidentally this sounds like a replication of a study James Adams[20] conducted some years ago when he was on leave from

your school district and was working with me at Ohio State University. An update of that study should provide useful direction for expansion of local district evaluation programs. Also, I am supportive of the state's reported plan to fund several local school districts to develop exemplary evaluation systems.

The fifth recommendation is that the Department abandon every pupil testing until there is clear cause for it. Again the state response is equivocal. In the text of their report the state staff agreed with the recommendation. But in the appendix they pointedly defended the continued testing of all pupils in selected subject fields, and at selected grade levels. The argument is that the cost is low and the benefit high. However, the state staff acknowledged the desirability of moving a matrix sampling plan for much of their assessment, and they announced an intent to encourage local districts to develop their own assessment programs.

Overall the state staff and I are not far apart on this recommendation. I still see no reason for testing every pupil on all test items, except for those school districts that want it and will use it. I do see the need for expanding the scope of state level assessment which would be feasible under a matrix sampling plan. And I believe that assisting local districts to strengthen their own assessment programs is consistent with the goal that I have already endorsed of improving evaluation capabilities at the local district level.

The sixth recommendation cited from our report by the state staff was that the Department validate its assessment tests with minority children. The state staff rejected this recommendation, because they said schools with high concentrations of minority children were heavily represented in the test tryouts. I find this a valid point that puts the need for a special validity study for minority children more in the direction of a desirable step than an absolutely essential one.

However, the staff response, in my view, evidences insensitivity to the validity problems of testing subgroups. Many years ago William Asher[21] demonstrated that "hill children from Kentucky" could be shown to have significantly higher I. Q.s than children from Columbus, Ohio, and that exactly the opposite result could be achieved by changing forms of the I. Q. test. The significant variable was not the test content but the language in which the content was conveyed. In my own experience, some years ago when I was participating in a large study in Texas my colleagues and I discovered a number of zero I. Q.s yielded from respectable I. Q. tests for which much validity information existed. Of course, the explanation was that Spanish-speaking students had been tested in English. Hopefully these two examples make my point: children, irrespective of their abilities, may expected to do poorly on tests that reflect language and cultural patterns that are foreign to them. For this reason it is not trivial to urge that the state staff conduct studies to determine whether their new tests are fair measure of reading and arithmetic for the minority group children in Michigan. If by no other means this question might be investigated through graduate student thesis research.

The seventh recommendation -- <u>that the Department encourage development of locally developed objectives</u> -- was endorsed by the state staff. At the same time the state staff reiterated their commitment to search for a set of common minimal skill objectives that would be mandated for all districts. While I am glad that the state staff will encourage and even assist the development of local objectives, I once again caution that the state should not label a set of objectives minimal and consensual until and unless they have obtained sound evidence to support such a claim.

In response to the eighth recommendation -- that the Department move the assessment program to matrix sampling -- the Department agreed. According to their response matrix sampling is considered an essential ingredient in future assessment designs. At the same time the Department sustained their commitment to continue every-pupil testing in reading, mathematics, and selected other skill areas. A part of the state's rationale for this is that a majority of fourth grade teachers surveyed by MEA[22] said they desire pupil-level information.

My response is that it is not necessary to compel all teachers to have their pupils tested on full-length tests just because over half the teachers desire this. Instead, such testing could be provided on a voluntary basis to those schools and teachers that desire the service. On balance, however, I am glad to see that the state will gear their program to matrix sampling, for this will make possible an expansion of the scope of variables that can be included in the state assessment program.

The ninth and final recommendation that the state staff referenced was to provide assistance and encouragement to local educators in the implementation of the accountability model. Not surprisingly, the state staff readily endorsed this recommendation. Further, they described their past efforts in this regard and expressed a commitment to continue and expand these efforts. Herein I believe the state has a unique opportunity to experiment with alternative forms of the six-step model and even with alternative accountability models. As they exploit these opportunities I am confident that the state department and school districts of Michigan will continue to provide national leadership in the area of educational accountability.

This concludes my discussion of the state's response to nine of the

recommendations that appeared in the House/Rivers/Stufflebeam report. By John Porter's count the state endorsed six and rejected three of them. By my count the state endorsed four, rejected two, and left three in doubt, since for those, the textual and appendix responses differed.

However, this is only part of the story since the nine recommendations addressed by the state staff were not the only ones included in our report. We also recommended: 10) that the common goals be periodically reviewed and updated; 11) that the Department put the objective-referenced testing program on a voluntary, service basis; and 12) that the Department expand the scope of its needs assessment work. As the state staff continues to plan and develop their accountability program I hope that they will consider these last three recommendations as well as the other nine that they have so far discussed.

Responses to Criticisms of the Panel Report

Beyond their reactions to the recommendations contained in our panel report, the state staff also criticized our work. They noted that ". . . to some degree, the panel succeeded in conducting an unbiased and objective evaluation of . . . Michigan's educational accountability program;" but that the report ". . . does contain inaccuracies, does not seem to be totally unbiased, and appears to be based on somewhat unrigorous and hurriedly-gathered information."

With one exception the charges of inaccuracy are invalid. We knew there were thirteen grade level panels that developed the Michigan objectives and did not say otherwise. We knew that many more objectives than those used for the objectives-referenced tests had been developed; but that point is moot since the objectives in use were not chosen randomly from the larger pool -- questions of scope and representativeness of the

objectives thus remain. We did study the Department's technical report on the objectives-referenced tests as implied in our comments about the good internal consistency within test item groups; we even obtained and studied the CTB/McGraw-Hill report[23] -- which raised serious questions about both the objectives and the items of the new tests. Although the Michigan Department of Education had contracted with CTB/McGraw-Hill for technical assistance in developing the new objectives-referenced tests, the latest information I have is that the Department has not accepted, made public nor paid for the CTB/McGraw-Hill report. On the contrary, in my view the Department suppressed this very timely and important report. Regarding another charge of inaccuracy, if the Chapter 3 program is not performance contracting, as we said, then why did John Porter characterize it in the N. Y. Times[24], as a "performance pact?" Finally, I agree that our panel was mistaken in referencing work on the norm-referenced tests in relation to our criticisms of the objectives-referenced tests.

At an earlier point in this paper I discussed in some detail the topic of the state staff's second criticism: objectivity. House, Rivers, and I anticipated and armed ourselves against charges of bias by obtaining advance written agreements that guaranteed our independence in conducting and reporting our study. Also, as our past work clearly shows, we are not prejudiced against the practice of evaluation and accountability. Thus, I reject -- with strong grounds -- the state's charges of bias and hope they will take note of the working agreements that governed the panel's work.

In response to their questions about rigor I would remind the state staff that our study was not a heavily funded long-term research project. It was a professional evaluation involving the study of source documents,

interviewing of personnel who represent a variety of interests in the Michigan system, the taking of oral and written testimony from many persons and groups, and the writing of a joint report. Each panelist devoted about eight days to the planning and implementation of the effort. We believe we collected important information, raised significant questions, proposed thoughtful recommendations, and communicated effectively with our intended audience. But I also encourage that others assess the merit of our report.

In my view our work should be judged for its balance in meeting criteria of technical adequacy, utility, impact, and cost/efficiency. Against these standards, our study was not "exhaustive" as characterized in the Kappan, is not the "last word on state accountability systems" as Helen Wise inferred, and does not depend solely for its adequacy on criteria for judging research as Kearney has implied. In general our study was intended to do what John Porter has since acknowledged it has done -- help ". . . focus attention and understanding on . . . the issue of educational accountability."

Concluding Remarks

Michigan has been a leader among the states in its development and operation of a statewide accountability system. Since it was started, the Michigan Accountability System has had a stormy history; it has been attacked from many sectors, especially within the state; and the state accountability staff have often counterattacked and then moved forward, or sideways, with continuing efforts to expand, improve and institutionalize their accountability work. As a consequence we, in Michigan, find ourselves a part of an ever expanding educational accountability movement.

Since this movement can affect the education of millions, the jobs of thousands, and the relationships between government and education, it behooves professional educators to be participants in the movement. We need to understand it, contribute to its growth, criticize it when we believe it to be errant, defend it when it is in the best interests of education in the state, and execute our professional responsibilities in carrying it out. Without such professional collaboration the accountability effort will surely fail, and the state will have wasted millions of dollars.

Those who are seeking support for the destruction of educational evaluation and accountability systems in Michigan and elsewhere should not claim they've found it in this report or in that by House, Rivers and Stufflebeam. Our works have been critical, but in the hope of pointing toward the improvement and not the demise of evaluation and accountability work. As our panel implied, everyone makes mistakes, and a good way of correcting them is by listening to what others have to say. In this respect I urge that MEA and the Michigan Department of Education continue to communicate and collaborate in their efforts to make education in Michigan a more responsive and accountable process.

As a final word I would remind the Michigan Department of Education staff, Michigan educators, and evaluators of accountability of some advice that Michael Scriven offered evaluators in his 1967, AERA tape[25]. I think it is particularly apropos to both the operation and evaluation of accountability systems.

> If you get into the kind of work illustrated in this case, you've got to face the fact from the beginning that your error rate is going to be high and the errors are going to be more visible than if you were working on rating proposals on a panel. If you can't stand the heat stay out of the kitchen. But on the

other side is the fact that in the kitchen you may be able to do something humanly valuable as well as educationally valuable and you might learn a good deal more. And if it isn't done by you it may be done worse.

References

1. Fred T. Wilhelms, (Chairman and Editor) ASCD 1967 Yearbook Committee. Evaluation as Feedback and Guide. p. 30.

2. Ernest House, Wendell Rivers, and Daniel Stufflebeam. An Assessment of the Michigan Accountability System. MEA/NEA, March 1974.

3. Technical Report for An Assessment of the Michigan Accountability System, Compiled by David Nevo, MEA, June 1974.

4. Stanley M. Elam (Editor), "Holding the Accountability Movement Accountable," Phi Delta Kappan, June 1974. pp. 657 and 674.

5. Helen Wise, On an Assessment of the Michigan Accountability System. Statement for Release to Press, NEA, April 8, 1974.

6. John Porter, Speech on educational accountability presented at the Annual Conference of the Open Court Press, La Salle, Illinois, 1974. (As reported in Education USA, June 5, 1974).

7. A Staff Response to the Report: An Assessment of the Michigan Accountability System. Michigan Department of Education, May 1974.

8. Edwin P. Bettinghaus and Gerald R. Miller, A Dissemination System for State Accountability Programs, Part I: Reactions to State Accountability Programs. Denver, Colorado: Cooperative Accountability Project, 1973.

9. The Ohio Accountability Project. Three Advocate Team Reports presented to the Ohio State Department of Education by the Ohio State University Evaluation Center, December 1972.

10. Ernest House, Wendell Rivers, Daniel Stufflebeam, "Assessment of Michigan's Accountability System," Teacher's Voice, Supplement 3. April 22, 1974.

11. House, Rivers, and Stufflebeam. "An Assessment of the Michigan Accountability System," Phi Delta Kappan, June 1974. pp. 663-669.

12. Technical Report, op. cit.

13. "A Description and Evaluation of Chapter 3 State Compensatory Education Programs in Michigan 1972-73," Michigan Department of Education, August, 1974.

14. Memo from John W. Porter to Members of the State Board of Education, State Compensatory Education--A Proposal for Revising Chapter 3 of the State School Aid Act for 1975-76, May 28, 1974.

15. "Michigan Compensatory Education Program Hailed," New York Times, December 11, 1973.

16. See for example "Officials Blast MEA Report," by William Cote, Kalamazoo Gazette, Saturday, May 4, 1974.

17. Staff Response, op. cit.

18. Memo from Robert Huyser to Superintendents and Assessment Coordinators of Local and Intermediate School Districts, Re: A Statewide Summary of Year 05 Assessment Results, March 11, 1974.

19. Memo from Frank B. Womer to Superintendents of Michigan school districts, Re: Interpretation of Michigan Educational Assessment Results, (Bureau of School Services, University of Michigan) January 1974.

20. James A. Adams, "A Study of the Status, Scope and Nature of Educational Evaluation in Michigan's k-12 School Districts." Ph.D. Dissertation, The Ohio State University, 1971.

21. E.J. Asher, "The Inadequacy of Current Intelligence Tests for Testing Kentucky Mountain Children." Journal of Genetic Psychology, 1935, 46, pp. 480-486.

22. Findings and Recommendations, Interim Report, NEA Task Force on Testing, May 29, 1973.

23. CTB Report, 1972-1973 Test Development and Experimental Tryout, Technical Report for the Michigan Department of Education, April 1973.

24. New York Times, op. cit.

25. Michael Scriven, "Evaluation Skills," Tape 6B. Produced by American Educational Research Association, May 1971.

Additional copies of papers in this series are available upon request for a fee of $2.00. Previous papers in this series are listed below. Please direct your inquiries to

Mary Anne Bunda, Series Editor
The Evaluation Center
College of Education
Western Michigan University
Kalamazoo, Michigan 49008

1. A Response to the Michigan Education Department's Defense of Their Accountability System by Daniel L. Stufflebeam.
2. An Administrator's Perspective of Evaluation by Jack P. Taylor.
3. Meta-Evaluation by Daniel L. Stufflebeam.
4. Evaluation Bias and Its Control by Michael Scriven.
5. Program Evaluation, Particularly Responsive Evaluation by Robert E. Stake.
6. A Basis for Determining the Adequacy of Evaluation Designs by James R. Sanders and Dean H. Nafziger.
7. Needed: Instruments as Good as Our Eyes by Henry M. Brickell.
8. Assessing the Impact of Planned Social Change by Donald T. Campbell.
9. A Checklist for Evaluating Large-Scale Assessment Programs by Lorrie A. Shepard.
10. Standards and Criteria by Gene V Glass.
11. Performance Standards by Nancy W. Burton.